

CUADERNOS DEL CIMBAGE



Universidad de Buenos Aires
Facultad de Ciencias Económicas



ESTUDIO DEL GASTO EN TURISTAS DE CRUCEROS EN URUGUAY PARA LA TEMPORADA 2010-2011 MEDIANTE EL ANÁLISIS DE REDES

Autor(es): Álvarez-Vaz, Ramón y Altmark, Silvia

Fuente: Cuadernos del CIMBAGE, Nº 21 (Junio, 2019), pp. 27 - 64

Publicado por: Facultad de Ciencias Económicas de la Universidad de Buenos Aires

Vínculo: <http://ojs.econ.uba.ar/ojs/index.php/CIMBAGE/issue/view/198>



Esta revista está protegida bajo una licencia Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Copia de la licencia: <https://creativecommons.org/licenses/by-nc-sa/4.0/>



Cuadernos del CIMBAGE es una revista académica semestral editada por el Centro de Investigaciones en Metodologías Básicas y Aplicadas a la Gestión (CIMBAGE) perteneciente al Instituto de Investigaciones en Administración, Contabilidad y Métodos Cuantitativos para la Gestión (IADCOM).

ESTUDIO DEL GASTO EN TURISTAS DE CRUCEROS EN URUGUAY PARA LA TEMPORADA 2010-2011 MEDIANTE EL ANÁLISIS DE REDES

Ramón Álvarez-Vaz*, Silvia Altmark**

Instituto de Estadística. Departamento de Métodos Cuantitativos.
Facultad de Ciencias Económicas y de Administración, Universidad de
la República. Eduardo Acevedo 1139, Montevideo CP11200, Uruguay

*ramon@iesta.edu.uy; ** salt@iesta.edu.uy

Recibido 22 de septiembre de 2018, aceptado 20 de marzo 2019

Resumen

El turismo de cruceros en Uruguay ha crecido desde la temporada 2004-2005, determinando un importante aporte de divisas cada temporada (abril a octubre de cada año). Por tanto, se estima relevante caracterizar las variables económicas involucradas en esta actividad, en particular, el gasto. A partir de los datos del Ministerio de Turismo de cruceros de la temporada 2010-2011, este trabajo compara la tipología surgida de aplicar métodos de clusters jerárquicos y no jerárquicos, con la que surge al aplicar el análisis de redes (SNA) a los datos del gasto en cruceiros. Previamente los autores habían construido una tipología de cruceiros, resultante de aplicar el algoritmo de Ward sobre distancias para variables binarias (gasta o no gasta) en cinco rubros de gasto. Los datos de cruceros se obtienen de una muestra de pasajeros a través de una encuesta cara a cara con diseño muestral complejo. Se cuenta con información de personas y gasto desde la temporada 2005-2006 y el presente trabajo se focaliza en la temporada 2010-2011. Para evaluar la performance del análisis de redes se trabaja con cuatro cruceros, (seleccionados al azar por probabilidad proporcional al aforo de cada uno) sobre los que, a partir de los gastos binarios, se construyen grafos, a los cuales se les aplican las diferentes métricas para su descripción. Usando la tipología previa de gastos, se estudia la asociación de las características socio-demográficas de los cruceiros con los grupos creados y con las comunidades identificadas con el SNA, para determinar eventuales patrones de comportamiento al cambiar de tipo de cruceiro.

Palabras clave: análisis de redes, clustering, gastos de cruceiros, métricas

Códigos JEL: C10, C30, C50

STUDY OF EXPENDITURE ON CRUISE TOURISTS IN URUGUAY FOR THE 2010-2011 SEASON USING NETWORK ANALYSIS

Ramón Álvarez-Vaz*, Silvia Altmark**

Instituto de Estadística. Departamento de Métodos Cuantitativos.
Facultad de Ciencias Económicas y de Administración, Universidad de
la República. Eduardo Acevedo 1139, Montevideo CP11200, Uruguay

*ramon@iesta.edu.uy; ** salt@iesta.edu.uy

Received September 22nd 2018, accepted March 20th 2019

Abstract

Cruise tourism in Uruguay has grown since the season 2004-2005, determining an important contribution of foreign currency each season (April to October of each year).

Therefore, it is considered relevant to characterize the economic variables involved in this activity, particularly expenditure.

Using data from Ministerio de Turismo (Ministry of Tourism) of cruise ships of the 2010-2011 season, this work compares the typology that emerged applying hierarchical and non-hierarchical clustering methods, with those emerged when applying the network analysis (SNA) to the expenditure data in cruise passengers.

Previously the authors had built a typology of cruise passengers, resulting from applying Ward's algorithm distances for binary variables (spend or not spend) in five items of expenditure.

Cruise data is obtained from a sample of passengers at through a face-to-face survey with complex sample design.

Information on people and spending is available since the 2005-2006 season; the present work focuses on the 2010-2011 season.

On the binary expenses in each cruise, graphs are constructed, to which several metrics are applied for its description.

Using the previous typology of expenditures, the authors study the association between socio-demographic characteristics of the cruise passengers with the groups created previously and the communities identified with the SNA, to determine patterns of behavior considering the expense when changing the type of cruise.

Keywords: Network Analysis, Clustering, Cruise Expenses, Metrics

JEL Codes: C10, C30, C50

1. INTRODUCCIÓN

En el marco de un crecimiento sostenido de la economía uruguaya desde hace catorce años, el turismo se debe destacar como una actividad de gran relevancia, en términos de divisas, valor agregado (PIB) y empleo.

El turismo en Uruguay es el primer sector exportador considerado individualmente. A su vez, de acuerdo a la información provista por el Ministerio de Turismo según la estimación de la Cuenta Satélite de Turismo, la actividad turística significó desde el año 2005 entre el 5 y el 8 % del PIB del país, siendo en 2018 el 8,6%. En 2017 generó 114.287 puestos de trabajo, que representan el 6,3% del conjunto de puestos totales de la economía - cifra que entre 2008 y 2017 se ha mantenido entre el 5 y el 6% -, según información obtenida en base a la Encuesta Continua de Hogares del Instituto Nacional de Estadística (INE).

Esta actividad genera otros impactos positivos en la economía, usufructuados por la población local: creación de infraestructuras y servicios, mejora de recursos humanos, aplicación de nuevas tecnologías, surgimiento de nuevas oportunidades de negocios, recuperación y/o preservación del patrimonio, puesta en valor de recursos. Debe advertirse, sin embargo, que la actividad turística también puede afectar negativamente un destino, cuando no se toman en consideración los impactos medioambientales y socioculturales de ciertas actividades. El Ministerio de Turismo de Uruguay ha desarrollado una mirada estratégica sobre la actividad, intentando mitigar este tipo de efectos.

En el caso del turismo de cruceros, además de su interés económico, se señala que es muy habitual que pasajeros de cruceros (“cruceiristas”), regresen al destino visitado, ya no en esta modalidad, sino como turistas, es decir, alojándose y permaneciendo algunos días en el país, con el consiguiente impacto positivo en la exportación de servicios turísticos. Debe tenerse en cuenta que el turismo de cruceros se trabaja por temporada, la cual en Uruguay se extiende, generalmente, entre octubre de un año y abril del siguiente.

De acuerdo a los datos relevados por el Ministerio de Turismo a partir de sus encuestas, el turismo de cruceros presenta una evolución creciente en Uruguay, partiendo de 75 arribos y 56.167 pasajeros desembarcados en la temporada 2004-2005 (desde cuando se dispone de datos). Entre esa temporada y la última (2017-2018), la cantidad de personas desembarcadas aumentó más del 300%. El máximo de arribos se da en 2013-2014, con 237 barcos, el mayor gasto se registra en 2012-2013 con 411.937 dólares corrientes y es en la temporada 2009-2010 donde se verifica el mayor gasto per cápita, U\$S 61.

Este desarrollo ha determinado la construcción de una nueva terminal de cruceros en el Puerto de Montevideo por parte de la Administración Nacional de Puertos, a lo que se agrega al hecho que desde diciembre de 2011 Montevideo es puerto de embarque y desembarque de pasajeros a cruceros, lo que implica un mayor derrame en la economía.

El siguiente cuadro muestra los cruceros arribados a los dos únicos puertos que reciben cruceros en Uruguay, Montevideo (MVD) y Punta del Este (PDE).

TEMPORADA	MVD	PDE	TOTAL	CRECIMIENTO ANUAL	
				MVD	PDE
2004-2005	63	12	75		
2005-2006	65	34	99	32%	183%
2006-2007	80	50	130	31%	47%
2007-2008	98	68	166	28%	36%
2008-2009	79	86	165	-1%	26%
2009-2010	83	96	179	8%	12%
2010-2011	76	95	171	-4%	-1%
2011-2012	119	106	225	32%	12%
2012-2013	122	101	223	-1%	-5%
2013-2014	121	116	237	6%	15%
2014-2015	108	64	172	-27%	-45%
2015-2016	105	56	161	-6%	-13%
2016-2017	98	52	150	-7%	-7%
2017-2018	86	54	140	-7%	4%

Tabla N°1. Cruceros arribados según puerto

Fuente: Área de Investigación y Estadística, Ministerio de Turismo

Es importante considerar las personas desembarcadas y su gasto, lo cual se evidencia en el cuadro siguiente:

TEMPORADA	GASTO	PERSONAS	GASTO/PERSONA
2004-2005	S/D	56.167	-----
2005-2006	4.241.639	110.827	38
2006-2007	11.235.466	149.062	75
2007-2008	16.818.273	256.593	66
2008-2009	14.384.413	247.120	58
2009-2010	17.830.909	292.048	61
2010-2011	13.291.304	278.627	48
2011-2012	20.884.091	353.727	59
2012-2013	18.612.467	411.937	45
2013-2014	18.855.505	409.371	46
2014-2015	10.943.470	332.118	33
2015-2016	11.141.587	317.205	35
2016-2017	9.798.264	260.704	38
2017-2018	7.692.437	242.466	32

Tabla N° 2. Gasto total de cruceristas

Fuente: Área de Investigación y Estadística, Ministerio de Turismo

La nacionalidad de los cruceristas que llegan a Uruguay es mayoritariamente brasileña, seguida de argentinos y luego norteamericanos y europeos; sin embargo atendiendo al gasto per cápita, son los europeos y norteamericanos quienes realizan el mayor gasto. En general el rubro de mayor relevancia en el gasto ha sido el de Compras, seguido de Alimentación. Las Compras se han visto fomentadas por algunas políticas de beneficios para los turistas, como el régimen de devolución de IVA (Tax Free), implantado desde diciembre de 2012 en Uruguay.

Teniendo en cuenta estos antecedentes y partiendo de trabajo anteriores, sobre los que se aplicaron técnicas de clustering (Álvarez Vaz et al., 2015) para crear tipologías de cruceristas, se propone complementar esos resultados a través del análisis de redes.

El presente trabajo se centra en la temporada 2010-2011, durante la cual llegaron 171 barcos (95 de ellos a Punta del Este), desembarcando 278.627 personas, que realizaron un gasto en dólares corrientes de 13.291.304.

El objetivo principal de este trabajo es caracterizar los grupos de cruceristas que viajan en los diferentes cruceros seleccionados y los objetivos específicos son los siguientes:

- Generar una red por cada tipo de crucero seleccionado y caracterizarla mediante diferentes métricas.
- Describir las características de cada red mediante el comportamiento de algunos nodos.
- Caracterizar los nodos más importantes de cada red analizada.
- Analizar si existe un patrón o estructura que se mantiene al cambiar de tipo de crucero (tamaño del crucero).

El trabajo está organizado de la siguiente forma: en la sección 2 se detalla la metodología; en la 3 se presenta el problema en estudio y se desarrollan las técnicas a aplicar y los datos que se utilizan; la sección 4 incluye los resultados, que se discuten en la sección 5; la sección 6, es donde se presentan las conclusiones y futuros pasos para la investigación en el tema.

2. MARCO TEÓRICO Y MÉTRICAS PARA CARACTERIZAR NODOS Y VÉRTICES

En esta sección se presentan las diferentes métricas que se usan para la caracterización de las redes. Para la presentación de las mismas se seguirá la notación de del libro 'Statistical Analysis of Network Data with R' (Kolaczyk y Csárdi, 2014), (Luke, 2015) aunque textos seminales como (Wasserman y Faust, 1994), (Borgatti et al., 2013) son una guía también a seguir.

GRADOS DE LOS VÉRTICES

Los grados de d_v de un vértice v de un grafo $G(V, E)$ es el número de aristas incidentes sobre V . A partir de esta medida se puede definir f_d como la fracción de vértices de $v \in V$ con grado $d_v = d$. El conjunto $\{f_d\}_{d \leq 0}$ es lo que se llama distribución de grados de G . Para las redes ponderadas, una generalización útil del grado es la noción de Fuerza de vértice que se obtiene simplemente sumando los pesos de los bordes de un vértice dado.

CENTRALIDAD DE LOS VÉRTICES

Las medidas de centralidad de intermediación tienen por objeto resumir en qué medida un vértice se encuentra 'entre' otros pares de vértices (Freeman, 1979) (*Betweenness centrality*)

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t|v)}{\sigma(s, t)}$$

donde $\sigma(s, t|v)$ es el número total de caminos más cortos entre s y t que pasan a través de v , y $\sigma(s, t)$ es el número total de caminos más cortos entre s y t (independientemente de si pasan o no por v). Esta medida de centralidad puede rescalarse al intervalo $[0, 1]$ mediante un factor de $(N_v - 1)(N_v - 2) / 2$, siendo N_v el número de vértices del grafo $G(V, E)$. Las medidas de centralidad de proximidad intentan capturar la noción de que un vértice es 'Central' si está 'cerca' de muchos otros vértices. (Freeman, 1979), (Brandes, 2001) El enfoque estándar, introducido por (Sabidussi, 1966), es dejar que la centralidad varíe inversamente con una medida de la distancia total de un vértice de todos los demás (*Closeness centrality*)

$$c_{cl}(v) = \frac{1}{\sum_{u \in V} dist(vb, u)}$$

Donde $dist(v, u)$ es la distancia geodésica entre los vértices $u, v \in V$. También para comparar entre otras medidas de centralidad, esta medida se puede rescalar al intervalo $[0, 1]$, a través de la multiplicación por un factor $N_v - 1$.

Finalmente, otras medidas de centralidad se basan en nociones de prestigio o rango. Es decir, buscan capturar la idea de que cuanto más centrales sean los vecinos de un vértice, más central es el vértice en sí.

mismo. Estas medidas pueden expresarse en términos de vectores propios de soluciones de sistemas lineales de ecuaciones y hay muchas medidas de centralidad de vectores propios.

De acuerdo a (Bonacich, 1987), (Bonacich y Lloyd, 2001)

$$C_{E_i} = \alpha \sum_{\{u,v\} \in E} C_{E_i}(u)$$

El vector $C_{E_i} = (C_{E_i}(1), \dots, C_{E_i}(N))^T$ es la solución al autovalor para $AC_{E_i} = \alpha^{-1} C_{E_i}$, donde A es la matriz de adyacencia para el grafo $G(V, E)$.

Bonacich sostiene que una elección óptima de $\alpha - 1$ es el autovalor más grande de A , y por lo tanto C_{E_i} es el autovector correspondiente. Cuando G es no dirigido el valor propio más alto de A será simple y su autovector tendrá valores distintos de cero y del mismo signo.

DESCRIPCIÓN DE LOS ENLACES

Se puede extender la idea de intermediación a los enlaces, aspecto que se denomina (*Edge betweenness centrality*) y que es una extensión de la intermediación de nodos asignando a cada enlace un valor que refleja el número de caminos más cortos (shortest paths), que atraviesan ese enlace. Para otras medidas de centralidad que caractericen los enlaces se puede consultar a (Brandes y Erlebach, 2005).

COHESIÓN DE LA RED

Existen varias maneras de evaluar la cohesión de una red, dependiendo del problema, donde puede usarse triadas o componentes gigantes así como lo que se denomina *cliques*, que no son más que subconjuntos de nodos totalmente cohesivos, en el sentido de que todos los vértices dentro del subconjunto están conectados por enlaces. Se pueden definir de tamaño 1 (que en este caso son los nodos v) mientras que cliques de tamaño 2 representan los enlaces (e); los cliques o subgrafos de tamaño 3 son lo que también (Kolaczyk y Csárdi, 2014) denominan *triangles*, de manera que al ir aumentando el tamaño de los cliques puede ver cuál es la estructura del grafo bajo análisis. Por el proceso de construcción antes descrito al aumentar el tamaño del clique, los últimos contienen los niveles más bajos, por lo cual (Kolaczyk y Csárdi, 2014) definen un concepto de clique máximo, que surge de considerar un subgrafo que no es subconjunto de un clique mayor y que da una métrica extra que los autores denominan *clique number*, que corresponde al tamaño del clique máximo.

DENSIDAD Y OTRAS MEDIDAS DE FRECUENCIA

Las caracterizaciones de la cohesión de red que surgen de los subgrafos y cliques consisten en ver con qué frecuencia y donde se encuentran a través de métricas como las que se presentan a continuación.

La densidad de un Gráfico es la frecuencia de los enlaces realizados en relación con el potencial número efectivo de bordes. Por ejemplo, en un grafo (no dirigido) G sin auto-bucles y sin múltiples enlaces, la densidad de un subgrafo

$$den(H) = \frac{|E(H)|}{|V(H)||V(H) - 1| * 0.5}$$

donde el valor de $den(H)$ estará entre $[0,1]$ y proporciona una medida de que tan cerca está el grafo H de ser un clique; En el caso de que G sea un grafo dirigido, el denominador en la ecuación (4) se sustituye por $|V(H)||V(H) - 1|$.

Otro concepto que se maneja para evaluar la densidad de la red es lo que (Kolaczyk y Csárdi, 2014) denominan *clustering coefficient*, donde surge un coeficiente de agrupamiento que puede medirse como

$$cl|T(G) = \frac{3\tau_{\Delta}(G)}{\tau_3(G)}$$

donde $\tau_{\Delta}(G)$ es el número de triángulos en el grafo G , y $\tau_3(G)$, el número de tripletas conectadas un subgrafo de tres vértices conectados por dos bordes, también A veces llamado un 2 estrellas). El valor $cl|T(G)$ se denomina alternativamente transitividad y es una métrica estándar en el estudio de las redes sociales, también conocido como “fracción de tripletas transitivas”. Por otra parte $cl|T(G)$ es una medida de la agrupación global o de clustering, resumiendo la frecuencia relativa con la que las tripletas conectadas forman triángulos cerrados.

CONECTIVIDAD

Una noción de conectividad es la que tiene que ver con el hecho de que si dado un subconjunto de k vértices (o enlaces) se quitan del grafo, el subgrafo restante aún permanece conectado. En particular un grafo $G(V,E)$ se llama k -vértice-conectado si el número de vértices $N_v > k$, y al eliminar cualquier subconjunto de vértices $X \in V$ de cardinalidad $|X| < k$ deja un subgrafo conectado.

A su vez, si $G(V,E)$ se denomina k -borde-conectado si $N_v \leq 2$, y al eliminar cualquier subconjunto de aristas $Y \in E$ de cardinalidad $|Y| < K$ deja un subgrafo que está conectado’. De esa manera se define como

conectividad de vértice (enlace) de $G(V, E)$ al entero más grande tal que G es k -vértice- (k -borde) conectado. (Kolaczyk y Csárdi, 2014) manifiestan que se puede demostrar que la conectividad del vértice está acotada por la conectividad de enlace, la que a su vez está acotada por el grado mínimo d entre los vértices en G .

CLUSTERING DE LA RED

Cuando se habla de partición de la red $C = \{C_1, C_2, \dots, C_k\}$, de un conjunto S se refiere a la división de la misma en clase naturales tales que S estas son disjuntas entre sí y a su vez la unión de ellas reproducen el conjunto de partida ($\bigcup_{k=1}^K C_k = S$). Pero, a su vez, es importante también evaluar si un subconjunto de nodos (algunas de esas clases) es 'cohesivo', para lo cual se entiende que es así si los nodos están bien conectados entre sí y, al mismo tiempo, están relativamente bien separados de los nodos restantes.

Así los algoritmos de particionado buscan una partición $C = \{C_1, C_2, \dots, C_k\}$, de un grafo $G = (V, E)$, de manera que los conjuntos $E(C_k, C_{k'})$ de enlaces conectando nodos de C_k en $C_{k'}$, sea relativamente pequeña en comparación al conjunto $E(C_k) = E(C_k, C_k)$ de enlaces que conectan nodos al interior de C_k .

Una primera forma de evaluar el particionado de la red es a través de clustering jerárquico, de tipo aglomerativo, donde se incorpora una función de costo, que refleja la cohesión, con lo cual surge el concepto de modularidad de C , donde se define $f_{kk}(C)$ como la fracción de enlaces de la red original que conectan nodos de C_i con nodos de C_j .

$$mod(c) = \sum_{k=1}^K [f_{kk}(C) - f_{kk}^*]^2$$

donde f_{kk} es el valor esperado de f bajo el supuesto de un modelo aleatorio de asignación de enlaces. Valores grandes de la modularidad sugieren que C captura una estructura no trivial de grupos (es decir que existen grupos), a la inversa si los enlaces se asignasen al azar.

ENLACE SELECTIVO (ASORTATIVIDAD)

Otra aspecto importante para evaluar la topología de una red es la evaluación de lo que se denomina enlace selectivo entre nodos de acuerdo a algunas características y que se miden con lo que se conoce como *Assortativity coefficient* y que tiene una lógica muy similar a la de los coeficientes de correlación. Este concepto a veces también se conoce como *homofilia*, y expresa la tendencia de las personas a relacionarse con personas que se le parecen.

Cuando la característica que se estudia es de tipo categórico (nominal u ordinal) la medida es:

$$r_a = \frac{\sum_i f_{ii} - \sum_i f_{i+} f_{+i}}{1 - \sum_i f_{i+} f_{+i}}$$

Donde f_{ii} es la fracción de enlaces en $G(V, E)$ que unen un nodo en la i -ésima categoría con un nodo en la j -ésima categoría y f_{+i}, f_{i+} expresan la suma de la i -ésima fila y columna respectivamente, de la matriz resultante f de frecuencias (Newman, 2002), (Newman, 2003).

El coeficiente descrito en la ecuación (8) está acotado en el intervalo $[-1, 1]$, expresando que si es cercano a 0, la mezcla de nodos en el grafo no difiere de la que se obtendría al asignar los enlaces al azar, preservando la distribución de grados marginal; cuando el coeficiente se acerca a 1 o -1 existe una mezcla selectiva perfecta.

Cuando los nodos tienen una característica de interés que es continua, para evaluar la *homofilia*, se consideran como (x_e, y_e) los valores que toman los nodos enlazados por el enlace e , para lo cual se usa el coeficiente de correlación de Pearson de los pares (x_e, y_e)

$$r = \frac{\sum_{x,y} xy - (f_{xy} - f_{x+} f_{+y})}{\sigma_x \sigma_y}$$

3. DESCRIPCIÓN DEL PROBLEMA EN ESTUDIO

Se trabaja solamente con 2225 cruceristas que son los que hicieron algún gasto. Dado que las unidades de análisis son grupos de cruceristas (individuales o múltiples), se propone armar redes para cada crucero. Como los cruceros no vienen una sola vez en la temporada, es necesario seleccionar los cruceros de la Tabla de datos diferenciando para cuál de las veces que vino en la temporada corresponde la información. Hay cruceros de diferente tamaño en términos de pasajeros, por lo cual se elegirán cruceros de diferentes capacidades y sobre esos se construirán las redes usando como criterios de enlace la cantidad de gastos que comparten. Se trabaja con el software Gephi (Bastian et al., 2009) y el R (R Core Team, 2016), para el que se usan varias librerías (Butts, 2016), (Csardi y Nepusz, 2006), (Kolaczyk y Cs, 2017).

4. RESULTADOS

Se crea una variable corte que sirve para saber qué cantidad de los 2225 grupos de cruceristas encuestados corresponde a un crucero en algunas de las veces que vino en la temporada.

Luego se crea una variable compuesta entre código de crucero y corte (que indica el ordinal de llegada en la temporada) para poder tener en la Tabla de datos los cruceros cada vez que vienen y es sobre esta variable que se van a seleccionar de tres a cinco unidades (son cruceros en cada oportunidad) y sobre eso se armarán las redes de cruceristas. De esta manera se tiene una variable que dimensiona a los diferentes cruceros en cada venida y teniendo en cuenta que solo se consideran los cruceristas que tuvieron algún gasto.

La forma de interpretar esta nueva variable es la siguiente:

- los últimos 2 dígitos reflejan el orden en el que ese crucero llegó durante la temporada
- el primero o los 2 primeros dígitos refieren al código numérico adjudicado al crucero como forma de identificarlo

Se puede ver a continuación el total de cruceristas encuestados en cada crucero que llegó en la temporada y fue seleccionado en la primera etapa de muestreo

Por ejemplo el crucero 9 llegó 7 veces (en el ordinal 14, 16, 18, 37, 59, 62 y 63 de los 82 cruceros que llegaron) y en esas 7 oportunidades se encuestaron 27, 28, 30, 34, 31, 32, 21 cruceristas. Pueden existir varias formas de seleccionar cruceros para analizarlos bajo la lógica del SNA y que a su vez sean diferentes entre sí para poder compararlos.

Se opta por clasificar a los cruceros de acuerdo a su capacidad de pasajeros, lo cual estaría dando lugar a redes de cruceristas provenientes de cruceros pequeños, medianos y grandes, lo que puede ser una característica importante para la estructura interna de cada red.

Resumiendo la Tabla N° 3 se ve que hay cuatro tamaños de cruceros: menos de 1000, entre 1000 y 2000, 2000 a 3000 y 3000 o más pasajeros. De esta manera se propone elegir cuatro cruceros, uno en cada estrato

SELECCIÓN DE LOS CRUCEROS

De las nuevas unidades 'cruceos veces' clasificadas en cuatro estratos, de acuerdo a la capacidad de cada crucero, se selecciona mediante

MAS (muestro aleatorio simple) uno en cada estrato, con los que se construyen redes, una para cada crucero en la oportunidad que llegó. En la temporada 2010-2011 se muestrearon un total de 82 cruceros.

De la Tabla N° 3 se ve que aparece el crucero con el que se construirá la red g_1 y la etiqueta 4453, significa que es el crucero 44 que fue muestreado a mitad de temporada (llegó en el lugar 53 de los 82), mientras que el crucero para la red g_3 es el 49, en el que se aplicó la encuesta casi a comienzos de la temporada.

Para el resto del análisis de los resultados, para cada crucero seleccionado se crea una red que se denominará g_1 , g_2 , g_3 , g_4 .

estrato	N_h	Crucero seleccionado
menos de 1000 (1)	7	4453
de 1000 a 2000 (2)	22	1644
de 2000 a 3000(3)	40	4912
3000 o más (4)	13	5150
Total	82	

Tabla N° 3. Descripción de los estratos de cruceros

Crucero	n_j	Red	Etiqueta
$j=1$	23	G_1	4453
$j=2$	38	G_2	1644
$j=3$	34	G_3	4912
$j=4$	32	G_4	5150

Tabla N° 4. Descripción de los cruceros analizados

FORMAS DE VISUALIZACIÓN DE LAS REDES

Para visualizar las redes y poner de manifiesto con claridad la estructura que existe en las mismas es importante manejar un concepto que en la literatura del SNA se denomina *layout*.

Por ejemplo, si se considera una red de tamaño moderado con 50 nodos, podría usarse una grilla cuadrada de 10 por 10, donde ir posicionando cada nodo. El primer nodo en la red podría ir en cualquiera de las 100 posiciones, el segundo nodo en 99 posiciones, y

así sucesivamente, lo que da una cantidad muy grande de diferentes diseños de red posibles.

Sin embargo muchas de estas posibles configuraciones pueden dar lugar a representaciones gráficas muy confusas, donde van a haber muchos enlaces que se cruzan, con lo cual la idea es elegir un diseño que tenga una probabilidad más alta de ser visualmente aceptable, por lo cual hay necesidad de recurrir a diferentes algoritmos que busquen maximizar ese efecto (Kolaczyk y Csi, 2017), (Csardi y Nepusz, 2006).

- Minimizar los cruces de enlaces.
- Maximizar la simetría de la disposición de los nodos
- Minimizar la variabilidad de las longitudes de los enlaces
- Maximizar el ángulo entre los enlaces cuando cruzan o unen nodos
- Minimizar el espacio total utilizado para la visualización de la red

En el Gráfico 1 puede verse cuál es el resultado de disponer un layout de tipo aleatorio, donde se ven varios enlaces que se cortan (esto debido a que la red para el caso del crucero 1 es chica); además el uso de esta forma de visualizar hace que la disposición sea distinta cada vez que se gráfica, al ser aleatoria la disposición espacial de cada nodo (por ende de cada enlace) en la red

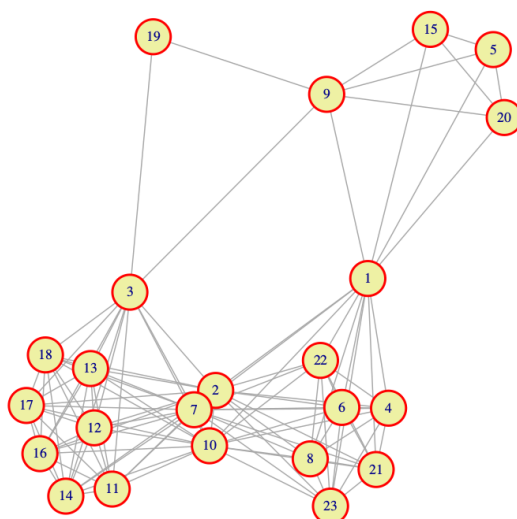


Gráfico N° 1. Red del crucero estrato 1 (layout aleatorio)

En este caso, como aparece en Gráfico 2, el diseño usado para visualizar el grafo, parte de ubicar el nodo 1 y disponer los restantes nodos en sentido trigonométrico negativo y donde de cada nodo parten los enlaces correspondientes. Este diseño tiene el inconveniente de generar grafos con mucha densidad de enlaces que se cruzan, sobre todo si el número de enlaces es muy grande y con lo cual no es claro el patrón de conexiones en la red

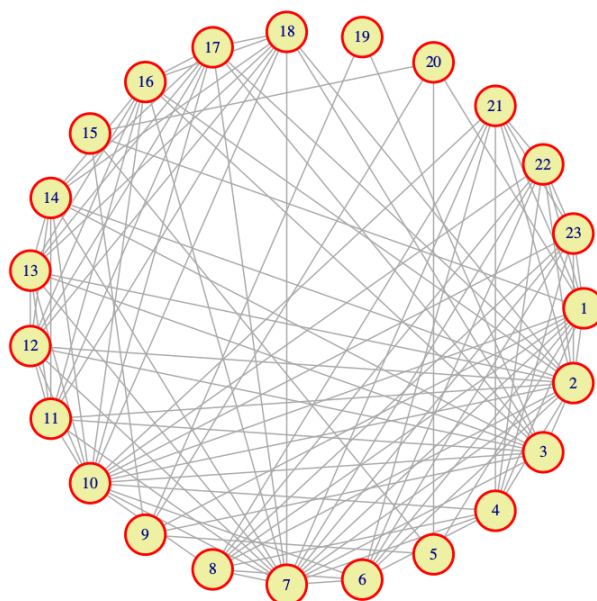
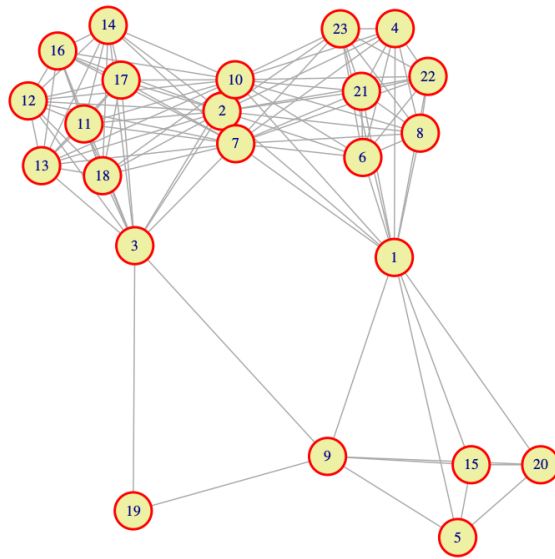


Gráfico N° 2. Red g1 del estrato 1 (*layout Circle*)

Se ha desarrollado un gran número de enfoques para el diseño automático de Gráficos de red. Una clase general de algoritmos, llamados de fuerza-dirigida, trabajan a partir de un enfoque en el que los nodos conectados tengan una fuerza de atracción similar a un resorte, al mismo tiempo que asignan fuerzas repulsivas a todos los pares de nodos. Los resortes en este algoritmo actúan para tirar de los nodos conectados más cerca uno del otro, mientras que las fuerzas repulsivas empujan los nodos no conectados lejos uno d otros. El sistema de red resultante se moverá y oscilará durante un tiempo, antes de llegar al estado estacionario que tiende a minimizar la energía en la red. Esto algoritmo, tiende a producir representaciones de las redes estéticamente agradables (Fruchterman y Reingold, 1991)



GráficoN° 3. Red g1 del estrato 1 (*layout Fruchterman-Reingold*)

RED DE CRUCERO 1 g_1

Los algoritmos usados en este caso para detectar comunidades estructurales tratan de encontrar subgrafos densos, usando paseos aleatorios (*random walk*) y donde la idea es que, los paseos aleatorios cortos tienden a estar en la misma comunidad (Pons y Latapy, 2015).

grados	frecuencia
4	1
6	3
8	1
11	6
12	7
14	1
15	19
19	3
$\overline{g_1} \cong 12$	Total=23

Tabla N° 5. Distribución de los grados de g1

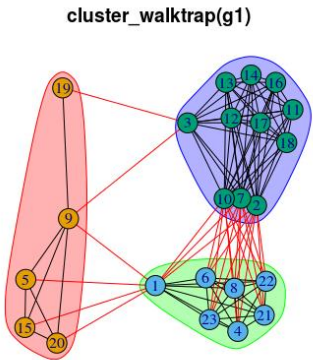


Gráfico N° 4. Clustering para Red g 1 del estrato 1 con algoritmo *walktrap*

Para poder determinar los clusters en la red se usa una función que tratar de evidenciar grafos densos, optimizando el score de

modularidad, a través del algoritmo fast greedy que aparece en Finding community structure in very large networks, <http://www.arxiv.org/abs/cond-mat/0408187> (Clauset et al., 2004).

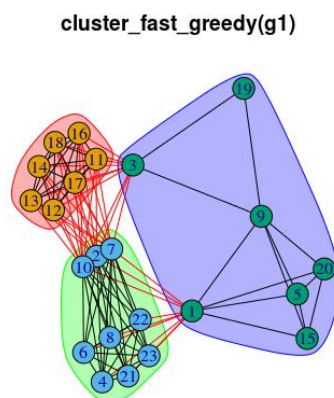


Gráfico N° 5. Clustering para Red g_1 del estrato 1 con algoritmo *fast greedy*

En este caso el criterio para determinar los clusters o comunidades basado en *edge-betweenness* es que, probablemente, los enlaces que conectan módulos separados tiene alto *edge-betweenness*, ya que los caminos más cortos de una comunidad a otra deben atravesarlos.

Por lo tanto, si gradualmente se van quitando enlaces con el más alto *edge betweenness*, se obtendría un mapa jerárquico que, en este caso, es un árbol raíz o *dendrograma*, donde los tallos representan los nodos individuales y la raíz el grafo enteramente.

El algoritmo *cluster-edge-betweenness* en forma iterativa calcula la intermediación de enlaces del grafo, quitando el enlace con más alta intermediación y volviendo a recalcular, hasta encontrar estabilidad (Newman y Girvan, 2004).

El algoritmo *label prop* es el más rápido, con performance de tiempo lineal, que se usa para detectar comunidades en redes. Comienza etiquetando los nodos con etiquetas únicas y luego adaptando las mismas por voto mayoritario en la vecindad (*neighborhood*) del nodo (Raghavan et al., 2007).

cluster_edge_betweenness(g1)

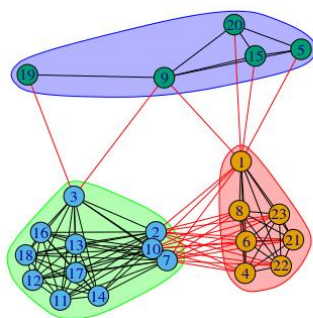


Gráfico N° 6. Clustering para Red g_1 del estrato 1 con algoritmo *edge betweenness*

cluster_label_prop(g1)

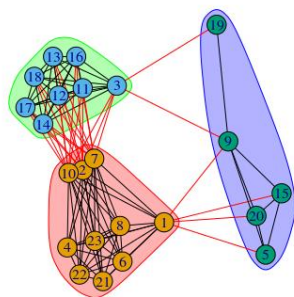


Gráfico N° 7. Clustering para Red g_1 del estrato 1 con algoritmo *label prop*

algoritmo	modularidad	Nro de clusters
cluster walktrap (cw.1)	0.2880	3
cluster fast greedy cfg.1)	0.2654	3
cluster edge etweenness (eb.1)	0.2880	3
cluster label prop (clp.1)	0.2923	3

Tabla N° 6. Modularidad para g_1 con diferentes algoritmos

CARACTERIZACIÓN DE LA RED

En la Tabla N° 7 se ven las características de los 23 nodos, donde se consignan las métricas de centralidad (cercanía e intermediación), grados de cada nodo, así como el número de triángulos que conforman cada uno de ellos.

En la Figura 8 se puede ver cómo es la relación entre el número de grados de cada nodo y la *betweenness*, donde resalta el nodo 1 que tiene 13 grados siendo el nodo más intermediario; en el Gráfico los puntos que aparecen con forma de estrella en rojo significa que en esa coordenada (valor de número de grado y *betweenness*) cae más de un nodo, por ejemplo para el valor 10 grados hay 7 nodos.

Para el caso de la cercanía se ve que la relación entre ésta y el número de grados es creciente, siendo los nodos 2 y 7 los más centrales y con mayor grado.

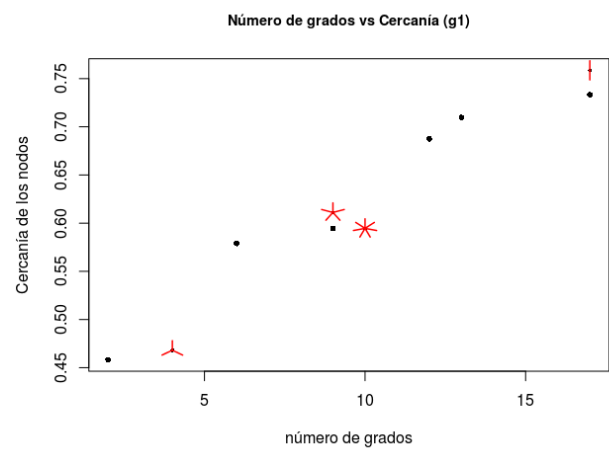


Gráfico N° 8. Relación entre grados y *betweenness* para Red g₁

Nodos	Grados	<i>Betweenness</i> (Intermediación)	<i>Closeness</i> (cercanía)	triángulos
1	13	0.23	0.71	42
2	17	0.12	0.76	80
3	12	0.12	0.69	46
4	9	0.00	0.31	36
5	4	0.00	0.47	6
6	9	0.00	0.61	36
7	17	0.12	0.76	80
8	9	0.00	0.59	36

Tabla N° 7 .Caracterización de los nodos de g₁

9	6	0.06	0.58	7
10	17	0.10	0.73	80
11	10	0.00	0.59	45
12	10	0.00	0.59	45
13	10	0.00	0.59	45
14	10	0.00	0.59	45
15	4	0.00	0.47	6
16	10	0.00	0.59	45
17	10	0.00	0.59	45
18	10	0.00	0.59	45
19	2	0.00	0.46	1
20	4	0.00	0.47	6
21	9	0.00	0.61	36
22	9	0.00	0.61	36
23	9	0.00	0.61	36

Tabla N° 8 . Caracterización de los nodos de g₁ –cont.-

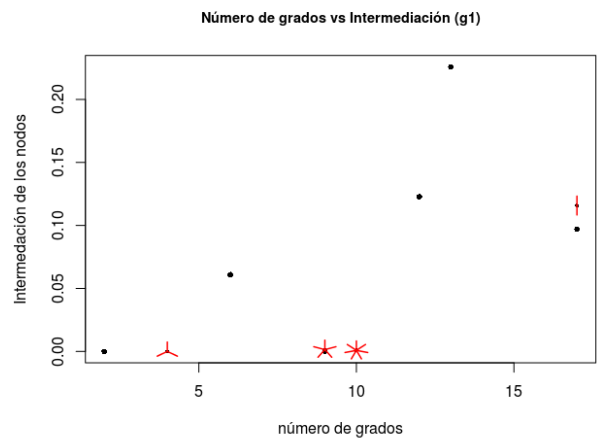


Gráfico N° 9. Relación entre grados y cercanía para Red g₁

	cw.1			cfg.1			eb.1			clp.1		
Nro de gastos	1	2	3	1	2	3	1	2	3	1	2	3
1	4	5	7	7	5	4	5	7	4	5	7	4
2	1	2	3	0	3	3	2	3	1	4	1	1
3	0	0	1	0	1	0	0	1	0	0	0	0

Tabla N° 9. Cluster (Número gastos) vs membresía para diferentes Algoritmos de clustering de redes: g₁

En la Tabla N° 8 puede verse como es la relación entre el número de gastos por cruceristas y la configuración de clusters que queda para g₁ usando los diferentes algoritmos de detección de comunidades presentados antes.

	<i>cw.1</i>			<i>cfg.1</i>			<i>eb.1</i>			<i>clp.1</i>		
<i>Cluster.bin</i>	1	2	3	1	2	3	1	2	3	1	2	3
1	4	6	10	7	7	6	6	10	4	8	8	4
2	0	1	1	0	2	0	1	1	0	2	0	0
3	1	0	0	0	0	1	0	0	1	0	0	1

Tabla N° 10 .Cluster (Cluster.Bin) vs membresía a diferentes algoritmos de clustering redes: g_1

En la Tabla N° 9 puede verse como es la relación entre los grupos creados mediante las técnicas de clustering convencional y la configuración de clusters que queda para g_1 usando los diferentes algoritmos de detección de comunidades presentados antes.

Para poder comparar cómo es la eventual asociación entre el clustering generado por el análisis de redes y la tipología previa que se tenía de los cruceristas mediante técnicas de clustering convencionales, en este caso mediante algoritmos jerárquico de Ward, usando distancias binarias, se presentan las siguientes Tablas:

	<i>Cluster.bin</i>				Total
	1	2	3	4	
Base completa	134	1748	212	131	
Crucero 1	0	20	2	1	23
Crucero 2	2	34	1	1	38
Crucero 3	2	21	1	0	24
Crucero 4	0	21	7	4	32
Total	4	96	11	6	117

Tabla N° 11. Distribución de Cluster.Bin para totalidad de cruceros y para los cuatro cruceros

El cluster que en la totalidad de cruceros tiene más frecuencia es el 2, con 78,5 %, aspecto que se repite para los cuatro cruceros seleccionados para la construcción de las redes, donde el grupo 2 tiene una presencia mayoritaria, en general con un 82 %.

Nodo	Gasto Tour	Gasto Alim	Gasto Shopp	Gasto Resto	Gasto Transp
19	1	0	0	0	0
15	0	0	0	1	0
12	0	0	1	0	0
7	0	1	1	0	0
Total crucero	3	10	11	5	2

Tabla N° 12. Perfil de gasto para 4 nodos de g_1

Una forma de comparar las diferencias entre las redes es evaluar cómo funciona el nivel de homofilia. Para eso se seleccionan 2 atributos que son la cantidad de gastos que tiene cada nodo y a su vez el tramo de gasto (quintil de gasto) de ese nodo en algunos de los gastos:

Red	Tipo de variable	Coeficiente de asortatividad
g_1	Nro de gastos=1	-0.143
g_1	Quintil 1 de Gasto total	-0.056
g_2	Nro de gastos=1	-0.09
g_2	Quintil 1 de Gasto total	-0.05
g_3	Nro de gastos=1	-0.134
g_3	Quintil 1 de Gasto total	0.165
g_4	Nro de gastos=1	-0.15
g_4	Quintil 1 de Gasto total	-0.07

Tabla N° 13. Ejemplos de Coeficientes de asortatividad para las diferentes redes

Para el resto de redes g_2 , g_3 , g_4 solamente se reportan los resultados correspondientes a las métricas evaluadas sobre las mismas y que aparecen en la Tabla N° 13.

COMPARACIÓN DE LAS CUATRO REDES

En la Tabla N° 13 se puede ver un resumen de las diferentes métricas evaluadas para los 4 grafos y con qué software se hizo. Si bien solamente aparece reseñados los valores y el software aplicado, en general ambos permiten la evaluación de las mismas.

métricas	g ₁	g ₂	g ₃	g ₄	software
número de nodos	23	32	24	32	igraph
número de enlaces	110	364	157	291	igraph
R ₁ (Razón enlaces/nodos)	4.8	9.6	6.5	9.1	
grado medio	11.6	21.2	15.1	20,2	gephi
grado medio con pesos	12.6	22.6	17.3	24.8	gephi
betweenness	7.55	10.87	5.94	8.09	igraph
closeness	0.027	0.017	0.029	0.021	igraph
eigen-centrality	0.485	0.570	0.536	0.560	igraph
transitividad de los nodos	0.792	0.837	0.807	0.834	igraph
excentricidad máxima	3	3	3	3	igraph
edge-density	0.434	0.517	0.568	0.586	igraph
conectividad de los enlaces	2	1	1	4	igraph
diámetro	3	3	3	3	Gephi
radio	2	2	2	2	gephi
densidad	0.526	0.572	0.656	0.651	gephi
modularidad	0.294	0.272	0.205	0.118	gephi
componentes conexos	1	1	1	1	gephi
coef. Medio de clustering	0.83	0.892	0.854	0.845	gephi
long media del camino	1.672	1.566	1.464	1.429	gephi

Tabla N° 14. Resumen de métricas para g₁, g₂, g₃, g₄

5. DISCUSIÓN

En la sección 4.3 se hace una identificación de la topología del grafo g_1 , a través de la evaluación de las diferentes métricas que caracterizan a los nodos y los enlaces y a la estructura de la misma, con el estudio de la cantidad de comunidades y el particionado a través de valores propios. También se estudia y se compara el grafo g_1 con redes aleatorias o de pequeño mundo con los mismos parámetros $G(24; 110)$. Para los restantes grafos, se presentan solamente las métricas principales que se consignan en la Tabla N° 13, aunque el procedimiento es equivalente.

Si se analizan los valores de la Tabla N° 13, las métricas globales de los cuatro grafos muestran grado medio que crece con la cantidad de nodos, siendo g_1 el que tiene menor ratio enlace/nodo, menor grado medio y menor densidad de enlaces y mayor modularidad y menor densidad de los cuatro grafos. El grafo g_3 es el que tiene menor intermediación de centralidad. En términos de cercanía las cuatro redes muestran un comportamiento similar, así como de transitividad de nodos. También el nivel de clusterización de las redes es similar y se destaca g_1 por ser el que tiene mayor longitud media de camino y menor centralidad de valores propios. A su vez, todos los grafos presentan, como era de esperar, un único componente conexo.

Además se describen las características socio demográficas y por gasto de los cuatro clusters binarios creados mediante técnicas de clustering convencional, considerando la totalidad de los cruceristas y cruceros (2225) para luego evaluar para los cuatro grafos (uno por crucero), cómo son las comunidades.

- El cluster 1 tiene casi un 95% de cruceristas con un solo gasto, siendo el cluster 4 el que diversifica el gasto, al tener casi 47% de gasto en dos rubros y 23% de gasto en tres rubros.
- El cluster 1 se caracteriza por ser el único que tiene el 100% del gasto en el quintil 5, mientras que el cluster 2 escalona el gasto en los 4 primeros quintiles, con casi nula presencia de cruceristas que gasten en el quintil 5 (menos del 4%); el cluster 3 tiene fuerte presencia de gastos en los quintiles 4 y 5 en partes casi iguales y por último el cluster 4 se diferencia por no tener cruceristas en los quintiles 1 y 2.

- con respecto al porcentaje de hombres que hay en cada grupo de cruceristas, el cluster 1 es el que tiene menor porcentaje de viajeros hombres, mientras que el cluster 4 es el que muestra mayor presencia de cruceristas hombres.
- Para evaluar la edad de los grupos de cruceristas se logra identificar que el cluster 4 es el grupo más joven, con casi 20 % de sus integrantes menores a 65 años, mientras que el cluster 1 es el que tiene mayoría de viajeros adultos, con casi 75 % mayores a 65 años.
- En cuanto a la composición de los clusters en términos de número de pasajeros, el cluster 2 es el que más se diferencia de los otros tres, al ser el que tiene casi 16 % de sus integrantes que son viajeros solos, mientras que el cluster 4 es el que tiene mayoría de cruceristas que viajan en grupos de a tres o más personas.

Teniendo en cuenta esa tipología, creada como si los 2225 pasajeros no viajasen en diferentes cruceros a lo largo de la temporada, es importante cruzar la pertenencia de cada nodo (un grupo de cruceristas de cada red, que coincide con la venida de un crucero en un día de la temporada) con la comunidad identificada. Para esta caracterización se usa solamente el algoritmo de *fast greedy*, que busca maximizar la modularidad, como fuera dicho en la sección 4.3.

	comunidades	1	2	3
Cluster.Bin	2	7	7	6
	3	0	2	0
	4	0	0	1
	Total	7	9	7

Tabla N° 15. Relación entre grupos (*Cluster.Bin*) y Comunidades detectadas para g_1

Para g_1 se puede ver que hay tres comunidades repartidas en forma homogénea y que a la interna del cluster 2, que es el mayoritario, también se reparten en forma pareja

	comunidades	1	2
Cluster.Bin	1	0	2
	2	20	14
	3	1	0
	4	1	0
	Total	22	16

Tabla N° 16. Relación entre grupos (Cluster.Bin) y Comunidades detectadas para g_3

Para los grafos g_2 y g_3 , a diferencia de como quedaba la distribución por clusters, donde había un grupo que se llevaba el 90%, en este caso surgen dos comunidades.

	comunidades	1	2
Cluster.Bin	1	2	0
	2	11	10
	3	0	1
	Total	13	11

Tabla N° 17. Relación entre grupos (Cluster.Bin) y Comunidades detectadas para g_2

	comunidades	1	2	3
Cluster.Bin	2	10	7	4
	3	2	4	1
	4	0	0	4
	Total	12	11	9

Tabla N° 18. Relación entre grupos (*Cluster.Bin*) y Comunidades detectadas para g_4

Para el grafo g_4 , a diferencia de cómo quedaba la distribución por clusters, donde había un grupo que se llevaba el 90%, en este caso surgen dos comunidades.

En la Tabla N° 18 se pueden ver los perfiles en términos de gastos de los nodos más relevantes de cada grafo, donde los que se seleccionaron son aquellos que tienen número de enlaces, *closeness* y *betweenness*, en valores mínimos, medios y máximos.

Por lo tanto, para el caso del nodo 19 de grafo g_1 , se está ante un nodo casi aislado, de los pocos que gasta en Tour, mientras que el nodo 7, con mayor número en enlaces y *closeness*, que comparte el gasto en Alimentación y Shopping (Compras).

Para el grafo g_2 , los nodos que sobresalen son el 6, que se caracteriza por tener un alto nivel de centralidad, solo conectado con otro nodo, que gasta en Transporte, mientras que un nodo como el 28, que gasta en Alimentación y Shopping, tiene 32 grados y una centralidad mediana a pesar de no tener casi intermediación.

Para el grafo g_3 , los nodos que se identifican son el 24 que solo gasta en Transporte, también solo conectado con otro nodo y muy bajo valor de centralidad y el nodo 11, que tiene el mayor grado, con casi 0.75 de centralidad y que se caracteriza por tener gasto en Alimentación y Shopping.

En cuanto al grafo g_4 , los nodos que se distinguen son el 6, que tiene un solo gasto (en Shopping), con un número medio de grados, alta

centralidad y baja intermediación, en contraposición con el nodo 24, de cruceristas que gastan en Alimentación y Shopping, un número alto de grados y de centralidad y también de triángulos.

El rubro mayoritario de gastos en general, tal como se ve en la última línea de la Tabla N° 18, corresponde a Shopping, con casi la mitad del gasto total, seguido por Alimentación, que absorbe una tercera parte del mismo, siendo los restantes rubros minoritarios. Esta relación, sin embargo, no se mantiene a la interna de los cruceros seleccionados para hacer el análisis de SNA, lo que estaría indicando que hay una cierta estructura en el perfil de gasto, que varía al cambiar de crucero.

Cuando se cambia de crucero (de tamaño de crucero), si bien la presencia del cluster 2, que corresponde a la tipología previa que es mayoritaria, no siempre los nodos se alinean con esta partición, lo que estaría indicando que la variabilidad intra crucero condiciona la partición en comunidades, que no se corresponden con los grupos creados usando todos los cruceristas de la temporada.

Nodo	G.Tour	G.Alim	G.Shopp	G.Resto	G.Trans	Comunidad
19	1	0	0	0	0	3
15	0	0	0	1	0	3
12	0	0	1	0	0	1
7	0	1	1	0	0	2
T. Crucero 1	3	10	11	5	2	3 comunidades
6	0	0	0	0	1	1
13	0	0	0	0	0	1
15	0	1	0	0	0	1

Tabla N° 19. Identificación de nodos usado algoritmo *fast greedy* por
Total de Gastos por rubro

28	0	1	1	0	0	2
T. Crucero 2	0	0	0	0	0	2 comunidades
24	0	0	0	0	1	1
5	0	1	1	0	0	2
19	0	1	0	0	0	2
11	0	1	1	1	0	1
T. Crucero 3	0	12	14	6	20	2 comunidades
6	0	0	1	0	0	1
15	0	0	0	1	0	3
7	0	1	0	0	0	2
24	0	1	1	0	0	2
T.crucero 4	6	16	20	5	5	3 comunidades
Total	267	1037	1411	1210	154	3079

Tabla N° 20. Identificación de nodos usado algoritmo *fast greedy* por
Total de Gastos por rubro-cont.-

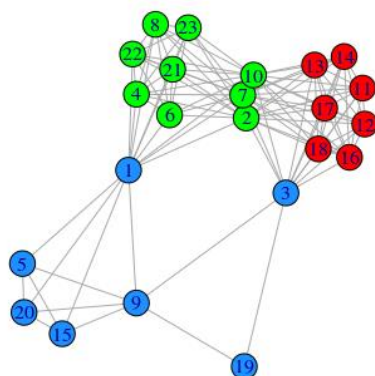


Gráfico N° 10. Clustering para Red g_1 contrastado de acuerdo a las comunidades formadas mediante algoritmo *fast greedy*

6. CONCLUSIONES

En forma preliminar, puede decirse que las cuatro redes, luego de ser caracterizadas en términos de las métricas de nodos y enlaces, muestran comportamientos diferentes entre sí. También las comunidades detectadas muestran una segmentación que difiere si se hace en forma global (clustering binario)

A su vez, en la segmentación creada con el SNA, aparecen comportamientos diferenciados en términos del nivel o número de gastos y que no se asocian directamente con la lógica usada para crear las redes en cada cruce, en base a si compartían alguno de los cinco rubros de gasto.

Por otra parte, en términos de la aplicación, se trabajó con el software Gephi, que tiene la gran ventaja de permitir visualizar grafos con muchos nodos y enlaces. Es una herramienta donde los gráficos son bastante parametrizables y que calcula una serie de métricas clásicas

en el SNA. Tiene la desventaja de ser una herramienta básicamente interactiva, con lo cual es necesario que el usuario haga los cambios personalmente.

En cambio el R tiene la ventaja de poder generar incluso más cantidad de métricas y, sobre todo, de ser un entorno de programación, donde se pueden incluso programar mediante funciones, otras métricas, simular diferentes tipos de redes y trabajar en SNA desde una perspectiva del modelado y la inferencia. Se aprovecha la potencia que tiene el R como lenguaje de programación orientado a objetos, que permite el tratamiento de éstos con diferentes bibliotecas desarrolladas en R. Sin embargo, la performance del R para la visualización de grafos, cuando estos son de tamaño mediano, es lenta.

Teniendo en cuenta los comentarios previos, se propone continuar la investigación en las siguientes líneas:

- Tomar algún estrato de tamaño de crucero y ampliar la muestra, creando de esa manera más redes y evaluando si se mantiene el patrón detectado con una sola.
- Crear una tipología con la lógica de clustering convencional (clasificación no supervisada) solamente sobre los nodos de la red y compararla con las comunidades de la red.
- Cambiar la forma de muestreo: en lugar de MAS, hacerlo proporcional al tamaño de cada crucero.
- Evaluar si un crucero que repite en la temporada, genera un grafo diferente y si puede depender del puerto de desembarque.
- Siendo lo planteado en el trabajo solamente descriptivo, proponer un análisis de redes validando modelos estadísticos, donde algunos de los atributos evaluados en la caracterización se puedan usar como variables explicativas, usando la teoría de los modelos exponenciales aleatorios en grafos (ERGM), (Kolaczyk y Csardi, 2014).
- Analizar el uso de los expansores en la lógica del SNA, ya que los datos provienen de una muestra con diseño muestral complejo, donde cada nodo (crucero encuestado) tiene un factor de expansión que incide en dos aspectos clave: para hacer inferencia y análisis factorial, ya que la inercia a

descomponer depende de las distancias de los nodos al baricentro y de los pesos de los mismos (Kolaczyk, 2009), (Kolaczyk y Csárdi, 2014).

BIBLIOGRAFÍA

- Álvarez Vaz, R., Altmark, S., y Santiñaque, F. (2015). Caracterización de los componentes del gasto de los turistas de cruceros en Uruguay mediante técnicas de datamining. En II Jornadas de Econometría, Facultad de Ciencias Económicas. UBA.
- Bastian, M., Heymann, S., y Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, (5):1170.
- Bonacich, P. y Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191 { 201.
- Borgatti, S. P., Everett, M. G., y Johnson, J. (2013). *Analyzing Social Networks*. SAGE Publications Ltd.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163{177.
- Brandes, U. y Erlebach, T. (2005). *Network analysis: methodological foundations*. Número 3418 en LCNS, Tutorial. Springer, Berlin ; New York. OCLC: ocm58474176.
- Butts, C. T. (2016). *sna: Tools for Social Network Analysis*. R package version 2.4.
- Clauset, A., Newman, M. E. J., y Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 0:066111.
- Csardi, G. y Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Fiedler, M. (1973). Algebraic connectivity of graphs. *Czech. Math. J*, 23(98):298{305.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215.
- Fruchterman, T. M. J. y Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129{1164.
- Kolaczyk, E. (2009). *Statistical analysis of network data : methods and models*. Springer, New York London.
- Kolaczyk, E. y Cs_ardi, G. (2014). *Statistical analysis of network data with R*. Springer, New York.

- Kolaczyk, E. D. y Cs, G. (2017). *sn: Statistical Analysis of Network Data with R*. R package version 1.0.3.
- Luke, D. (2015). *A user's guide to network analysis in R*. Springer, Cham.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.
- Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.
- Newman, M. E. J. y Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.
- Pons, P. y Latapy, M. (2015). Computing communities in large networks using random walks.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghavan, U. N., Albert, R., y Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E*, 76:036106.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581-603.
- Wasserman, S. y Faust, K. (1994). *Social network analysis: methods and applications*. Número 8 en *Structural analysis in the social sciences*. Cambridge University Press, Cambridge ; New York.